GenCore version 4.5

OM protein - protein search, using sw model

Run on: May 29, 2002, 11:38:48 ; Search time 13.47 Seconds
(without alignments)
1859.803 Million cell updates/sec

Title: US-08-485-355B-50
Perfect score: 3374
Sequence: 1 MGDAGVASQRPHNRRGTRNV.....GKIAARVRARRARRRAARAN 647

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	130	3.9	708	1	VP40_HCMVA	P16753 human cytom
2	123	3.6	886	1	VGP3_EBVA8	Q07284 epstein-bar
3	121.5	3.6	907	1	VGP3_EBV	P03200 epstein-bar
4	119	3.5	1363	1	ILPR_BRALA	O02466 branchiosto
5	115.5	3.4	699	1	CHI1_BACCI	P20533 bacillus ci
6	115	3.4	1419	1	ALA1_CANAL	O13368 candida alb
7	113.5	3.4	781	1	COAT_PAVHB	P07299 human parvo
8	113	3.3	969	1	SACB_STRSL	Q55242 streptococc
9	112	3.3	1090	1	GUXB_CELFI	P50899 cellulomona
10	111	3.3	1902	1	P1P_LACLC	P16271 lactococcus
11	110.5	3.3	505	1	GUNV_ERWCA	Q47096 erwinia car
12	109	3.2	591	1	FLGE_CAUCR	P35806 caulobacter
13	108.5	3.2	444	1	GUNN_ERWCA	Q59394 erwinia car
14	108	3.2	775	1	VP4_ROTH3	P39033 human rotav
15	108	3.2	1260	1	ALS1_CANAL	P46590 candida alb
16	107.5	3.2	2090	1	N214_HUMAN	P35658 homo sapien
17	106.5	3.2	609	1	HSF_SCHPO	Q02953 schizosacch
18	106.5	3.2	776	1	VP4_ROTB6	P36306 bovine rota
19	106.5	3.2	2201	1	POLG_CXA9	P21404 c genome po
20	106	3.1	1138	1	TIE1_HUMAN	P35590 homo sapien
21	105	3.1	775	1	VP4_ROTF1	P39034 feline rota
22	105	3.1	775	1	VP4_ROTHJ	Q01641 human rotav
23	104.5	3.1	670	1	VG50_HSVI1	Q00130 ictalurid h
24	104.5	3.1	1481	1	APU_THEET	P38939 t amylopull
25	104.5	3.1	2255	1	RRPL_SV5	Q88434 simian viru
26	104.5	3.1	2255	1	RRPL_SV5WR	Q03396 simian viru
27	103.5	3.1	1052	1	MGPC_MYCGE	P22747 mycoplasma
28	103.5	3.1	1223	1	YWR1_CAEEL	Q10925 caenorhabdi
29	103	3.1	1122	1	ADP1_MYCGA	Q49379 mycoplasma
30	102.5	3.0	776	1	VP4_ROTBU	P12474 bovine rota
31	102.5	3.0	2628	1	HAGA_PORGI	Q51845 porphyromon
32	102	3.0	666	1	MUR2_ENTHR	P39046 enterococcu
33	102	3.0	2029	1	LAR_DROME	P16621 drosophila
34	101	3.0	1216	1	NKX1_BOVIN	Q28139 bos taurus
35	101	3.0	2514	1	POLN_SINDO	P27283 sindbis vir
36	100.5	3.0	519	1	IRX4_HUMAN	P78413 homo sapien
37	100.5	3.0	562	1	NUPL_HUMAN	P52594 homo sapien
38	100.5	3.0	784	1	SX15_DROME	P40657 drosophila
39	100.5	3.0	1228	1	SLAP_BACST	P35825 bacillus st
40	100.5	3.0	2481	1	UN52_CAEEL	Q06561 caenorhabdi
41	100	3.0	504	1	GUNW_ERWCA	Q59395 erwinia car
42	100	3.0	555	1	EG10_CAEEL	P49809 caenorhabdi
43	100	3.0	1433	1	SUBF_BACSU	P16397 bacillus su
44	100	3.0	1625	1	CTPI_MYCTU	Q10900 mycobacteri
45	100	3.0	2832	1	NDVB_RHIME	P20471 rhizobium m

ALIGNMENTS

RESULT 1
VP40_HCMVA
ID VP40_HCMVA STANDARD; PRT; 708 AA.
AC P16753; Q69030;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Capsid protein P40 [Contains: Assemblin (Protease) (EC 3.4.21.97);
DE Capsid assembly protein].
GN UL80 OR APNG.
OS Human cytomegalovirus (strain AD169).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10360;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90269039; PubMed=2161319;
RA Chee M.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R.,
RA Horsnell T., Hutchison C.A. III, Kouzarides T., Martignetti J.A.,
RA Preddie E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.;
RT "Analysis of the protein-coding content of the sequence of human
RT cytomegalovirus strain AD169.";
RL Curr. Top. Microbiol. Immunol. 154:125-169(1990).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF PROTEASE.
RX MEDLINE=96399135; PubMed=8805706;
RA Tong L., Qian C., Massariol M.-J., Bonneau P.R., Cordingley M.G.,
RA Lagace L.;
RT "A new serine-protease fold revealed by the crystal structure of
RT human cytomegalovirus protease.";
RL Nature 383:272-275(1996).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF PROTEASE.
RX MEDLINE=96399136; PubMed=8805707;
RA Qiu X., Culp J.S., Dilella A.G., Hellmig B., Hoog S.S., Janson C.A.,
RA Smith W.W., Abdel-Meguid S.A.;
RT "Unique fold and active site in cytomegalovirus protease.";
RL Nature 383:275-279(1996).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.27 ANGSTROMS) OF PROTEASE.
RX MEDLINE=96399137; PubMed=8805708;
RA Shieh H.-S., Kurumbail R.G., Stevens A.M., Stegeman R.A.,
RA Sturman E.J., Pak J.Y., Wittwer A.J., Palmier M.O., Wiegand R.C.,
RA Holwerda B.C., Stallings W.C.;
RT "Three-dimensional structure of human cytomegalovirus protease.";
RL Nature 383:279-282(1996).
CC -!- FUNCTION: THE CAPSID ASSEMBLY PROTEIN IS A COMPONENT OF THE CAPSID
CC CORE INVOLVED IN PROCESSING AND PACKAGING OF PROGENY DNA.
CC ASSEMBLIN IS A PROTEASE WHICH CAN PROTEOLYTICALLY CLEAVE ITSELF
CC AND THE CAPSID ASSEMBLY PROTEIN AT THE C-TERMINUS.
CC -!- CATALYTIC ACTIVITY: PREFERENTIALLY CLEAVES AT ALA-|-SER OR ALA-|-
CC ALA.
CC -!- CATALYTIC ACTIVITY: CLEAVES -ALA + SER- AND -ALA + ALA- BONDS IN
CC THE SCAFFOLD PROTEIN.
CC -!- PTM: CAPSID ASSEMBLY PROTEIN IS PHOSPHORYLATED.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S21 (SERINE PROTEASE).

```
CC   -!- SIMILARITY: TO OTHER HERPESVIRUSES CAPSID PROTEIN VP40.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X17403; CAA35353.1; -.
DR   EMBL; X17403; CAA35354.1; ALT_INIT.
DR   PIR; S09843; QQBEB8.
DR   PDB; 1WPO; 15-OCT-97.
DR   PDB; 1LAY; 26-SEP-97.
DR   PDB; 1CMV; 04-SEP-97.
DR   MEROPS; S21.002; -.
DR   InterPro; IPR001847; Assemblin.
DR   Pfam; PF00716; Peptidase_S21; 1.
DR   PRINTS; PR00236; HSVCAPSIDP40.
KW   Coat protein; Hydrolase; Serine protease; Phosphorylation;
KW   3D-structure.
FT   CHAIN         1    256       ASSEMBLIN (PROTEASE).
FT   CHAIN       257    643       CAPSID ASSEMBLY PROTEIN (BY SIMILARITY).
FT   PROPEP      644    708       C-TERMINAL PEPTIDE (BY SIMILARITY).
FT   CHAIN         1    708       GENE UL80 PROTEIN.
FT   CHAIN       336    708       GENE UL80.5 PROTEIN.
FT   CHAIN       393    708       GENE UL80.4 PROTEIN.
FT   CHAIN       478    708       GENE UL80.3 PROTEIN.
FT   SITE        256    257       CLEAVAGE (BY THE PROTEASE).
FT   SITE        643    644       CLEAVAGE (BY THE PROTEASE) (PROBABLE).
FT   ACT_SITE     63     63       CHARGE RELAY SYSTEM.
FT   ACT_SITE    132    132       CHARGE RELAY SYSTEM.
FT   ACT_SITE    157    157       CHARGE RELAY SYSTEM.
SQ   SEQUENCE   708 AA;  73851 MW;  32A993D6586824C9 CRC64;
```

```
  Query Match             3.9%;  Score 130;  DB 1;  Length 708;
  Best Local Similarity   21.9%;  Pred. No. 0.13;
  Matches   89;  Conservative   42;  Mismatches  151;  Indels  124;  Gaps   15;

Qy    269 ALTPTSLPQYDVSEAYAL-HTLTFARPSSAAALAFVWAGLPQGGTAPAGTPAWEQASSGG 327
          | :| ||   : :|| :: | |: | |:: |         | | |    :||    ||
Db    319 AESPPSLSPSEPAEAASMSHPLSAAVPAATA---------PPGATVAGASPA---VSS-- 364

Qy    328 YLTWRHNGTTFPAGSVSYVLPEGFALERYDPNDGSWTDFASAGDTVTFRQVAVDEVVVTN 387
           | | |:|   |  :   :|
Db    365 -LAWPHDGVYLPKDAFFSLL---------------------------------------- 383

Qy    388 NPAGGGSAPTFTVRVPPSNAYTNTVFRNTLLETRPSSRRLELPMP--PADFGQTVANNPK 445
            | |    |  | : |              ||:    ||:|  || :|  |    :
Db    384 ----GASRSAVPVMYPGAVA------------APPSASPAPLPLPSYPASYGAPVVGYDQ 427

Qy    446 IEQSLLKETL-------GCYLVHSKMRNPVFQLTPASSFGAVSFNNPGYERTRDLPD--- 495
          :      : :       |     :   :| : : |  |         | | | || |
Db    428 LAARHFADYVDPHYPGWGRRYEPAPSLHPSYPVPPPPS--------PAYYRRRDSPGGMD 479

Qy    496 --------YTGIRDSFDQNM-------------STAVAHFRSLSHSCSIVTKTYQGWEG 533
                 | |      |                 || |   |     |   :   |
Db    480 EPPSGWERYDGGHRGQSQKQHRHGGSGGHNKRRKETAAASSSSSDEDLSFPGEAEHGRAR 539

Qy    534 ---VTNVNTPFGQFAHAGLLKNEEILCLADDLATRLTGVYPATDNFAAAVSAFAANMLSS 590
            ::||:  |   |||   |::     |:|   :  :    | |||  |   : :||:
Db    540 KRLKSHVNSDGGSGGHAG--SNQQQQQRYDELRDAIHEL--KRDLFAARQS---STLLSA 592

Qy    591 VLKSEATSSIIKSVGETAVGAAQSGLAKLPGLLMSVPGKIAARVRA 636
          | | |:||   :   | |   ||  : |  |:|   |:| | :|
Db    593 ALPSAASSSPTTTTVCTPTGELTSGGGETPTALLSGGAKVAERAQA 638
```

```
RESULT   2
VGP3_EBVA8
ID   VGP3_EBVA8     STANDARD;      PRT;   886 AA.
```

```
AC   Q07284;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Envelope glycoprotein GP340 (Membrane antigen) (MA).
GN   BLLF1.
OS   Epstein-barr virus (strain AG876) (Human herpesvirus 4), and
OS   Epstein-barr virus (strain P3HR-1) (Human herpesvirus 4).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Lymphocryptovirus.
OX   NCBI_TaxID=82830, 82829;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AG876, AND P3HR-1;
RX   MEDLINE=93331716; PubMed=8393237;
RA   Lees J.F., Arrand J.E., Pepper S.V., Stewart J.P., Mackett M.,
RA   Arrand J.R.;
RT   "The Epstein-Barr virus candidate vaccine antigen gp340/220 is highly
RT   conserved between virus types A and B.";
RL   Virology 195:578-586(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=P3HR-1;
RA   Klein K., Mueller-Lantzsch N.;
RL   Submitted (OCT-1992) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: RESPONSIBLE FOR EBV BINDING TO THE CR2 RECEPTOR ON HUMAN
CC       B-CELLS.
CC   -!- SUBCELLULAR LOCATION: MOST ABUNDANT COMPONENT OF THE VIRAL
CC       ENVELOPE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L07922; AAA02783.1; -.
DR   EMBL; L07923; AAA02787.1; -.
DR   EMBL; X67776; CAA47986.1; -.
KW   Membrane; Glycoprotein; Antigen; Late protein.
FT   CARBOHYD     47     47       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     87     87       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    114    114       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    166    166       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    169    169       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    195    195       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    229    229       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    277    277       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    318    318       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    328    328       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    345    345       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    356    356       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    378    378       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    386    386       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    411    411       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    435    435       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    443    443       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    457    457       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    497    497       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    519    519       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    533    533       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    568    568       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    589    589       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    624    624       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    680    680       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   886 AA;  92388 MW;  4394F6130DECCA8A CRC64;
```

```
  Query Match             3.6%;  Score 123;  DB 1;  Length 886;
  Best Local Similarity   20.3%;  Pred. No. 0.58;
  Matches  138;  Conservative   66;  Mismatches  269;  Indels  206;  Gaps   29;
```

```
Qy      48 NFTAAAQDLAQSLDANTVTFPANISSMPEFRNWAKGKIDLDSDSIGWYFKYLDPAGATES 107
           | ||  :  || ||   || | :: :  :  |::     : :: :|     :|     |
Db     169 NITAVVR--AQGLD---VTLPLSLPTSAQDSNFS-----VKTEMLG---NEID----IEC 211

Qy     108 ARAVGEYSKIPDGLVKFSVDAEIRE-------IYNEECPVVTDVSVPLDGRQWSLSIFSF 160
              || |::  |  ||::     |        |     || |   :|  |  :|| :
Db     212 IMEDGEISQVLPGDNKFNITCSGYESHVPSGGILTSTSPVAT--PIPGTGYAYSLRLTPR 269

Qy     161 PMFR------TAYVAVANVENKEMSLD-------VVNDLI----EWLNNLADWRYVVDSE 203
           |: |        ||  :    |    |       | :| |    :   |  |  || |:
Db     270 PVSRFLGNNSILYVFYSGNGPKASGGDYCIQSNIVFSDEIPASQDMPTNTTDITYVGDN- 328

Qy     204 QWINFTNDTTYYVRIRVLRPTYDVPDPTEGLVRTVSDYRLTYKAITCEANMPTLVDQGFW 263
                           || ||  |                    :|| | :    ||
Db     329 -------------------ATYSVPMVTSE----------------DANSPNVTVTAFW 352

Qy     264 --------------------------IGGQYALTPT----------------------- 273
                                       | | :|   |
Db     353 AWPNNTETDFKCKWTLTSGTPSGCENISGAFASNRTFDITVSGLGTAPKTLIITRTATNA 412

Qy     274 ---------SLPQYDVSEAYALHTLTFARPSSAAALAFVWAGLPQGGTAPAGTPAWEQAS 324
                  |      : :  |:|  || |::          |||     |    |   ||
Db     413 TTTTHKVIFSKAPESTTTSPTLNTTGFAAPNTT-------TGLPSSTHVPTNLTA--PAS 463

Qy     325 SGGYLTWRHNGTTFPAGSVSYVLPEGFALERYDPNDG-------SWTDFASAGDTVTFRQ 377
           :|  ::      :  |||: |   |         | |          |   ||  | |
Db     464 TGPTVSTADVTSPTPAGTTSGASP---VTPSPSPRDNGTESKAPDMTSPTSAVTTPTPNA 520

Qy     378 VAVDEVVVT-----NNPAGGGSAPTFTVRVPPSNAYTNT---------VFRNTLLETRPS 423
           :    | |      :|  | ::||  |  |  || : |             || :| |:
Db     521 TSPTPAVTTPTPNATSPTLGKTSPTSAVTTPTPNATSPTPAVTTPTPNATIPTLGKTSPT 580

Qy     424 SRRLELPMPPADFGQTVANNPKIEQSLLKETLGCYLVHSKMRNPVFQLTPASSFGAVSFN 483
           |  :  | | |         :|  : :    |||          ||    | ::  ||:
Db     581 S-AVTTPTPNATSPTVGETSP--QANTTNHTLG-----GTSSTPVVTSPPKNATSAVTTG 632

Qy     484 NPGYERTRDLPDYTGIR-DSFDQNMSTAVAHFRSLSHSCSIVTKTYQGWEGVTNVNTPFG 542
              :  |        :|  |  : :| : :   | ||   : :    | | :| | ||
Db     633 Q--HNITSSSTSSMSLRPSSISETLSPSTSD-NSTSHMPLLTSAHPTGGENITQV-TPAS 688

Qy     543 QFAHAGLLKNEEILCLADDLATRLTGVYPAT-------DNFAAAVSAFAANMLSSVLKSE 595
              |               ::|      | |        | : :      |:
Db     689 TSTH-------------HVSTSSPAPRPGTTSQASGPGNSSTSTKPGEVNVTKGTPPKN 734

Qy     596 ATSSIIKSVGETAVGAAQS 614
           |||    |  :|||   |
Db     735 ATSPQAPSGQKTAVPTVTS 753
```

RESULT 3
VGP3_EBV
ID VGP3_EBV STANDARD; PRT; 907 AA.
AC P03200; P03201;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Envelope glycoprotein GP340 (Membrane antigen) (MA) [Contains:
DE Glycoprotein GP220].
GN BLLF1.
OS Epstein-barr virus (strain B95-8) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10377;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84270667; PubMed=6087149;
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA Tuffnell P.S., Barrell B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL Nature 310:207-211(1984).
CC -!- FUNCTION: RESPONSIBLE FOR EBV BINDING TO THE CR2 RECEPTOR ON HUMAN
CC B-CELLS.
CC -!- SUBCELLULAR LOCATION: MOST ABUNDANT COMPONENT OF THE VIRAL
CC ENVELOPE.
CC ---

CC ---
DR EMBL; V01555; CAA24854.1; -.
DR PIR; A03762; QQBE21.
DR PIR; A03763; QQBE22.
DR PIR; S33008; S33008.
KW Membrane; Glycoprotein; Antigen; Late protein; Alternative splicing.

FT	CARBOHYD	47	47	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	87	87	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	114	114	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	166	166	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	169	169	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	195	195	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	229	229	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	277	277	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	318	318	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	328	328	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	345	345	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	356	356	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	378	378	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	386	386	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	411	411	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	435	435	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	443	443	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	457	457	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	497	497	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	519	519	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	533	533	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	547	547	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	568	568	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	589	589	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	610	610	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	624	624	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	627	627	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	645	645	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	656	656	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	683	683	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	701	701	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	735	735	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	746	746	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	755	755	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	780	780	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	815	815	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	858	858	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	888	888	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	502	698	MISSING (IN GP220).

SQ SEQUENCE 907 AA; 94431 MW; 0750141CBCAC52C9 CRC64;

Query Match 3.6%; Score 121.5; DB 1; Length 907;
Best Local Similarity 19.6%; Pred. No. 0.77;
Matches 138; Conservative 66; Mismatches 266; Indels 233; Gaps 28;

```
Qy      48 NFTAAAQDLAQSLDANTVTFPANISSMPEFRNWAKGKIDLDSDSIGWYFKYLDPAGATES 107
           | ||  :  || ||   || | :: :  :  |::     : :: :|     :|     |
Db     169 NITAVVR--AQGLD---VTLPLSLPTSAQDSNFS-----VKTEMLG---NEID----IEC 211

Qy     108 ARAVGEYSKIPDGLVKFSVDAEIRE-------IYNEECPVVTDVSVPLDGRQWSLSIFSF 160
              || |::  |  ||::     |        |     || |   :|  |  :|| :
Db     212 IMEDGEISQVLPGDNKFNITCSGYESHVPSGGILTSTSPVAT--PIPGTGYAYSLRLTPR 269

Qy     161 PMFR------TAYVAVANVENKEMSLD-------VVNDLI----EWLNNLADWRYVVDSE 203
           |: |        ||  :    |    |       | :| |    :   |  |  || |:
```

```
Db    270 PVSRFLGNNSILYVFYSGNGPKASGGDYCIQSNIVFSDEIPASQDMPTNTTDITYVGDN- 328

Qy    204 QWINFTNDTTYYVRIRVLRPTYDVPDPTEGLVRTVSDYRLTYKAITCEANMPTLVDQGFW 263

Db    329 ------------------ATYSVPMVTSE----------------DANSPNVTVTAFW 352

Qy    264 -------------------------IGGQYALTPT----------------------- 273

Db    353 AWPNNTETDFKCKWTLTSGTPSGCENISGAFASNRTFDITVSGLGTAPKTLIITRTATNA 412

Qy    274 ---------SLPQYDVSEAYALHTLTFARPSSAAALAFVWAGLPQGGTAPAGTPAWEQAS 324

Db    413 TTTTHKVIFSKAPESTTTSPTLNTTGFADPNTT-------TGLPSSTHVPTNLTA--PAS 463

Qy    325 SGGYLTWRHNGTTFPAGSVSYVLPEGFALERYDPNDGSW-----------TDFASAGDTV 373

Db    464 TGPTVSTADVTSPTPAGTTSGASP-------VTPSPSPWDNGTESKAPDMTSSTSPVTTP 516

Qy    374 TFRQVAVDEVVVT-------------------NNPAGGGSAPTFTVRVPPSNAYTNTVFR 414

Db    517 TPNATSPTPAVTTPTPNATSPTPAVTTPTPNATSPTLGKTSPTSAVTTPTPNATSPTLGK 576

Qy    415 N----------------TLLETRPSSRRLELPMPPADFGQTVANNPKIEQSLLKETLGCY 458

Db    577 TSPTSAVTTPTPNATSPTLGKTSPTS-AVTTPTPNAT-GPTVGETSP-QANATNHTLG-- 631

Qy    459 LVHSKMRNPVFQLTPASSFGAVSFNNPGYERTRDLPDYTGIRDSFDQNMSTAVAHFRSLS 518

Db    632 ---GTSPTPVVTSQPKNATSAVTTGQ--HNITSSSTSSMSLRPSSNPETLSPSTSDNSTS 686

Qy    519 HSCSIVTKTYQGWEGVTNVNTPFGQFAHAGLLKNEEILCLADDLATRLTGVYPAT----- 573

Db    687 HMPLLTSAHPTGGENITQV-TPASISTH-------------HVSTSSPAPRPGTTSQAS 731

Qy    574 --DNFAAAVSAFAANMLSSVLKSEATSSIIKSVGETAVGAAQS 614

Db    732 GPGNSSTSTKPGEVNVTKGTPPQNATSPQAPSGQKTAVPTVTS 774


RESULT   4
ILPR_BRALA
ID   ILPR_BRALA     STANDARD;      PRT;  1363 AA.
AC   O02466;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Insulin-like peptide receptor precursor (EC 2.7.1.112) (ILP receptor).
OS   Branchiostoma lanceolatum (Common lancelet) (Amphioxus).
OC   Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC   Branchiostoma.
OX   NCBI_TaxID=7740;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96408719; PubMed=8813726;
RA   Pashmforoush M., Chan S.J., Steiner D.F.;
RT   "Structure and expression of the insulin-like peptide receptor from
RT   amphioxus.";
RL   Mol. Endocrinol. 10:857-866(1996).
CC   -!- FUNCTION: THIS RECEPTOR BINDS TO THE INSULIN RELATED PEPTIDE AND
CC       HAS A TYROSINE-PROTEIN KINASE ACTIVITY.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: PROBABLE TETRAMER OF 2 ALPHA AND 2 BETA CHAINS LINKED BY
CC       DISULFIDE BONDS. THE ALPHA CHAINS CONTRIBUTE TO THE FORMATION OF
CC       THE LIGAND-BINDING DOMAIN, WHILE THE BETA CHAIN CARRY THE KINASE
CC       DOMAIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE INSULIN RECEPTOR FAMILY OF TYROSINE-
CC       PROTEIN KINASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; S83394; AAB50848.1; -.
DR   HSSP; P06213; 1IRK.
DR   InterPro; IPR000494; EGFR_L.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR002174; Furin-like.
DR   InterPro; IPR002011; Receptor_tyr_kin_II.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00041; fn3; 3.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF01030; Recep_L_domain; 1.
DR   PRINTS; PR00014; FNTYPEIII.
DR   PRINTS; PR00109; TYRKINASE.
DR   SMART; SM00060; FN3; 3.
DR   SMART; SM00261; FU; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS00239; RECEPTOR_TYR_KIN_II; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Tyrosine-protein kinase; Receptor; Transmembrane;
KW   Glycoprotein; ATP-binding; Phosphorylation; Signal.
FT   SIGNAL        1     29       POTENTIAL.
FT   CHAIN        30    716       INSULIN-LIKE PEPTIDE RECEPTOR, ALPHA-
FT                                 CHAIN (POTENTIAL).
FT   PROPEP      717    720       POTENTIAL.
FT   CHAIN       721   1363       INSULIN-LIKE PEPTIDE RECEPTOR, BETA-
FT                                 CHAIN (POTENTIAL).
FT   DOMAIN      721    928       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    929    949       POTENTIAL.
FT   DOMAIN      950   1363       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      994   1283       PROTEIN KINASE.
FT   NP_BIND    1000   1008       ATP (BY SIMILARITY).
FT   BINDING    1028   1028       ATP (BY SIMILARITY).
FT   ACT_SITE   1148   1148       BY SIMILARITY.
FT   MOD_RES    1174   1174       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CARBOHYD     51     51       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     97     97       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    137    137       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    278    278       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    599    599       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    617    617       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    665    665       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    666    666       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    711    711       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    732    732       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    736    736       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    743    743       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    816    816       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    885    885       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    898    898       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1363 AA;  154104 MW;  238120B4EAB1ED65 CRC64;


  Query Match             3.5%;  Score 119;  DB 1;  Length 1363;
  Best Local Similarity   19.8%;  Pred. No. 2.1;
  Matches   99;  Conservative   60;  Mismatches  162;  Indels  180;  Gaps   23;

Qy     79 NWAKGKIDLDSDSIGWYFK----------YLDPAGATESARAVGEYSKIPDGLVKFSVDA 128

Db    416 DWDRTDITIDEGKLFFHFNPKLCRHVILTMVDKVGLPEHA--------ITD--------- 458

Qy    129 EIREIYNEECPVVTDVSVPLDGRQWSLSIFSFPMFRTAYVAVANVENKEMSLDVVNDLIE 188

Db    459 ------------TDISTLTNGDQ------------AQCSFSRLE-----------IE 480
```

```
Qy    189 WLNNLADWRYVVDSEQWINFTNDTTYYVRIRVLRPTYDVPDPTEGLVRTVSDYRLTYKAI 248
Db    481 EINTSKDMIIL----RWSEF-------------RP----PDPRDLLSYTVS-YRET---- 514

Qy    249 TCEANMPTLVDQGF-WIGGQYALTPTSLPQYDVSEAYALHTLTFARPSSAAAL---AFVW 304
Db    515 ---------EDQGIDEYDGQDACGNTEWKEFDVSPTQTAHIITGLKPWTQYALLVKTYTK 565

Qy    305 AGLPQGG----------------TAPAGTPAWEQASSGGYLTWRHNGTTFPAGSVSYVLP 348
Db    566 AGAREGSGAKSDIVYARTDADKPTHPQDVVVYSNSSNTLIITWKPPNR--PNGNVTHYIV 623

Qy    349 EGFALERYDPNDGSWTDFASAG-----DTVTFRQVAVDEVVVTNNPAGGGSAPTFTVRVP 403
Db    624 K-YKRQQEDVAEMEQREYCKGGLKPHRPTQGLEDIVNNEEEPNNSTIGDGTC----CECP 678

Qy    404 PS-----------------NAYTNTVFRNTLLETRPSSRRLELPMPPADF--GQTV--- 440
Db    679 KSEDEIRIEEEEAAFQQEFENFLHNNVYHKRENETRAGRRRRELPVTARPFYSNQTVNVT 738

Qy    441 --------------ANNPKIEQSLLKETLGCYLVHSKMRN---PVFQLTPASSFGAVSFN 483
Db    739 LPSTNRTVPPTPTPNPNPQLETTVWNEHM---VVLTGLRHFSEYIIEVIACNADAAVGCS 795

Qy    484 NPGYERTRDLPDYTGIRDSFD 504
Db    796 GSAVELARTQAD-----DSAD 811
```

```
RESULT   5
CHI1_BACCI
ID   CHI1_BACCI     STANDARD;      PRT;   699 AA.
AC   P20533;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   01-FEB-1995 (Rel. 31, Last annotation update)
DE   Chitinase A1 precursor (EC 3.2.1.14).
GN   CHIA1.
OS   Bacillus circulans.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1397;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WL-12;
RX   MEDLINE=90368776; PubMed=2203782;
RA   Watanabe T., Suzuki K., Oyanagi W., Ohnishi K., Tanaka H.;
RT   "Gene cloning of chitinase A1 from Bacillus circulans WL-12 revealed
RT   its evolutionary relationship to Serratia chitinase and to the type
RT   III homology units of fibronectin.";
RL   J. Biol. Chem. 265:15659-15665(1990).
RN   [2]
RP   MUTAGENESIS.
RC   STRAIN=WL-12;
RX   MEDLINE=93366760; PubMed=8103047;
RA   Watanabe T., Kohori K., Miyashita K., Fujii T., Sakai H.,
RA   Uchida M., Tanaka H.;
RT   "Identification of glutamic acid 204 and aspartic acid 200 in
RT   chitinase A1 of Bacillus circulans WL-12 as essential residues for
RT   chitinase activity.";
RL   J. Biol. Chem. 268:18567-18572(1993).
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-
CC       acetyl-D-glucosamine polymers of chitin.
CC   -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M57601; AAA81528.1; -.
DR   PIR; A38368; A38368.
DR   HSSP; P07254; 1CTN.
DR   InterPro; IPR003610; Chitin_bind3.
DR   InterPro; IPR001579; Chitinase_2.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR001223; Glyco_hydro_18.
DR   Pfam; PF02839; CBD_7; 1.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF00704; Glyco_hydro_18; 1.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00495; ChtBD3; 1.
DR   SMART; SM00060; FN3; 2.
DR   PROSITE; PS01095; CHITINASE_18; 1.
KW   Hydrolase; Glycosidase; Chitin degradation; Signal; Repeat.
FT   SIGNAL       1     41
FT   CHAIN       42    699       CHITINASE A1.
FT   DOMAIN      42    460       CATALYTIC.
FT   DOMAIN     465    549       FIBRONECTIN TYPE-III (R-1).
FT   DOMAIN     560    644       FIBRONECTIN TYPE-III (R-2).
FT   ACT_SITE   204    204       PROTON DONOR (PROBABLE).
FT   MUTAGEN    200    200       D->N: DECREASE IN ACTIVITY.
FT   MUTAGEN    200    200       D->E: NO CHANGE IN ACTIVITY.
FT   MUTAGEN    204    204       E->D,Q: LOSS OF ACTIVITY.
SQ   SEQUENCE   699 AA;  73677 MW;  AC7C9B22E2987643 CRC64;
```

```
  Query Match             3.4%;  Score 115.5;  DB 1;  Length 699;
  Best Local Similarity   21.4%;  Pred. No. 1.4;
  Matches  114;  Conservative   54;  Mismatches  207;  Indels  157;  Gaps   25;

Qy    195 DWRYVV------DSEQWINFTNDTTYYVRIRVLRPTYDVPDPTEGLVRTVSDYRLTYKAI 248
Db    202 DWEYPVSGGLDGNSKRPEDKQNYTLLLSKIREKLDAAGAVDGKKYLLTIASGASATYAAN 261

Qy    249 TCEANMPTLVDQGFWIGGQYALTPTSLPQYDVSEAY---ALHTLTFARPSSAAALAFVWA 305
Db    262 TELAKIAAIVD---WI---------NIMTYDFNGAWQKISAHNAPLNYDPAASA-----A 304

Qy    306 GLPQGGT----------APAGTPA-------------WE---QASSGGYLTWRHNGT--T 337
Db    305 GVPDANTFNVAAGAQGHLDAGVPAAKLVLGVPFYGRGWDGCAQAGNGQYQTCTGGSSVGT 364

Qy    338 FPAGS-------VSYVLPEGFALERYDPNDGSWTD-------------FASAGDT---- 372
Db    365 WEAGSFDFYDLEANYINKNGYT--RY------WNDTAKVPYLYNASNKRFISYDDAESVG 416

Qy    373 -------------VTFRQVAVDEVVVTNN------PAGGGSAPTFTVRVP-PSNAYTNTV 412
Db    417 YKTAYIKSKGLGGAMFWELSGDRNKTLQNKLKADLPTGGTVPPVDTTAPSVPGNARSTGV 476

Qy    413 FRNTL-LETRPSSRRLELPMPPADFGQTVANNPKIEQSLLKETL--------GCYLVHSK 463
Db    477 TANSVTLAWNASTDNVGVT------GYNVYNGANLATSVTGTTATISGLTAGTSYTFTIK 530

Qy    464 MRNPVFQLTPASSFGAVSFNNPGYERTRDLPDYTGIRDSFDQNMSTAVAHFRSLSHSCSI 523
Db    531 AKDAAGNLSAASNAVTVS--------------TTAQPGGDTQAPTAPTNLASTAQTTSS 575

Qy    524 VTKTYQGWEGVTNVNTPFGQFAHAGLLKNEEILCLADDLATRLTGVYPATDNFAAAVS-A 582
Db    576 ITLS---WTASTDNVGVTGYDVYNGTA----------LATTVTGTTATISGLAADTSYT 621

Qy    583 FAANMLSSVLKSEATSSI--IKSVGET---AVGAAQSGLAKLPGLLMSVPGK 629
Db    622 FTVKAKDAAGNVSAASNAVSVKTAAETTNPGVSAWQVNTAYTAGQLVTYNGK 673
```

RESULT 6

```
ALA1_CANAL
ID   ALA1_CANAL     STANDARD;      PRT;  1419 AA.
AC   O13368;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Agglutinin-like protein ALA1 precursor (Agglutinin-like adhesin).
GN   ALA1 OR ALS5.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98053977; PubMed=9393828;
RA   Gaur N.K., Klotz S.A.;
RT   "Expression, cloning, and characterization of a Candida albicans
RT   gene, ALA1, that confers adherence properties upon Saccharomyces
RT   cerevisiae for extracellular matrix proteins.";
RL   Infect. Immun. 65:5289-5294(1997).
CC   -!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.
CC   -!- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF025429; AAB88883.1; -.
KW   Cell adhesion; Glycoprotein; Repeat; Signal.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18   1419       AGGLUTININ-LIKE PROTEIN ALA1.
FT   DOMAIN      399    404       POLY-THR.
FT   DOMAIN      408    418       POLY-THR.
FT   DOMAIN      437    441       POLY-THR.
FT   DOMAIN      673    676       POLY-SER.
FT   DOMAIN      687    690       POLY-SER.
FT   DOMAIN      700    703       POLY-SER.
FT   DOMAIN      719    724       POLY-SER.
FT   DOMAIN      749    752       POLY-SER.
FT   DOMAIN      787    791       POLY-SER.
FT   DOMAIN      869    872       POLY-SER.
FT   DOMAIN      875    883       POLY-SER.
FT   DOMAIN      901    911       POLY-SER.
FT   DOMAIN     1216   1221       POLY-SER.
FT   CARBOHYD    665    665       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    919    919       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1301   1301       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1326   1326       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1419 AA;  149635 MW;  249F33F688A9D5B6 CRC64;
```

```
  Query Match             3.4%;  Score 115;  DB 1;  Length 1419;
  Best Local Similarity   17.7%;  Pred. No. 4.3;
  Matches  123;  Conservative   90;  Mismatches  249;  Indels  232;  Gaps   26;

Qy      39 TGRQVSPPDN--FTAAAQDLAQSLDANTVTFPANISSMPEFRNWAKGKID---------- 86
           ||  |   |:  |||           |||||      :    |: |  :|
Db     139 TGSSVDLEDSKCFTAG---------TNTVTFNDGSKKLSIAVNFEKSTVDQSGYLTTSRF 189

Qy      87 -------------------LDSDSIGWYFKYLDPA------------------------ 102
                             | ::|:   | | |
Db     190 MPSLNKIATLYVAPQCENGYTSGTMGFSTSYGDVAIDCSNVHIGISKGVNDWNHPVTSES 249

Qy     103 -GATESARAVG---EYSKIPDGLVKFSVDAEIREIYNEECPVVTDVSVPLDGRQWSLSIF 158
            |:|  : |    |  :| |   | :|| |               | |  |: ||
Db     250 FSYTKSCSSFGISITYQNVPAGYRPF-IDAYIS--------------PSDNNQYQLS-- 291

Qy     159 SFPMFRTAYVAVANVENKEMSLDVVNDLIEWLNNLADWRYVVDSEQWINFTNDTTYYVRI 218
               ::  |  | :                        |::   : :|  : |       |
```

```
Db     292 ----YKNDYTCVDDY---------------------WQHAPFTLKWTGYKNSDAGSNGI 325

Qy     219 RVLRPTYDVPDPTEGLVRTVSDYRLTYKAITCE--ANMPTLVDQGFWIG---------- 265
           ::  |  | | |   | |:       |  | |    :||      ::|
Db     326 VIVATTRTVTDSTTA-VTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYSTKTAP 384

Qy     266 -GQYALTPTSLPQYD----VSEAYALHTLTFARPSSAAALAFVWAGLPQGGTAPAGTPAW 320
            |: |     :| :      ||     | |  | :   ::  |   :|       |  |
Db     385 IGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVPSPNPTTTTTQFW 444

Qy     321 EQASSGGYLTWRHNGTTFPAGSVSYVLPEGFALERYDPNDGS------WTDFASAGDTVT 374
           ::      |     |    |: | :: |        |:: :      |::  :  :|:|
Db     445 SES-----FTSTTTITNSLKGTDSVIVRE--------PHNPTVTTTEFWSESFATTETIT 491

Qy     375 FRQVAVDEVVV----------------------TNNPAGGGSAPTFTVRVPPSNAYTN 410
           :    | |:|                       || | |  |     || | :   |
Db     492 SKPEGTDSVIVREPHNPTVTTTEFWSESYATTETITNGPEGTDSV---IVREPHNPTVTT 548

Qy     411 TVFRNTLLETRPSSRRLELPMPPADFGQTVANNPK-IEQSLLKETLGCYLVHSKMRNPVF 469
           | | :    |              :|: | |:  :  ::||     :  :: :  :
Db     549 TKFWSESYATT---------------ETITNKPEGTDSVIVKEPYNPTVTTTEFWSESY 592

Qy     470 QLTPASSFG----------------AVSFNNPGYERTRDL------PDYTGIRDSFDQ 505
           |   : |                  | :  |  |  :       |   | |  ::
Db     593 ATTETITNGPEGTDSVIVREPHNPTVTTTEFWSESYATTETITTGPLGTDSIVIHDPLEE 652

Qy     506 NMSTAVAHFRSLSHSCSIVTKTYQGWEGVTNVNTPFGQFAHAGLLKNEEILCLADDLATR 565
           : ||      ::  | | :: : |  | ::|       :  ||  : :|   :|  ::
Db     653 SSSTT-----AIESSDSNISSSAQ--ESSSSVEQ---SSSIVGLSSSSDIPLSSDMPSSS 702

Qy     566 LTGVYPATDNFAAAVSAFAANMLSSVLKSEATSS 599
           ||:   | : :: ||:: ::  ||:  |  :||
Db     703 STGL---TSSESSTVSSYDSDSSSSIESSTLSSS 733
```

```
RESULT   7
COAT_PAVHB
ID   COAT_PAVHB     STANDARD;      PRT;   781 AA.
AC   P07299;
DT   01-APR-1988 (Rel. 07, Created)
DT   01-APR-1988 (Rel. 07, Last sequence update)
DT   01-APR-1990 (Rel. 14, Last annotation update)
DE   Probable coat protein VP1.
OS   Human parvovirus B19.
OC   Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Erythrovirus.
OX   NCBI_TaxID=10798;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ISOLATE AU;
RX   MEDLINE=86200451; PubMed=3701931;
RA   Shade R.O., Blundell M.C., Cotmore S.F., Tattersall P., Astell C.R.;
RT   "Nucleotide sequence and genome organization of human parvovirus B19
RT   isolated from the serum of a child during aplastic crisis.";
RL   J. Virol. 58:921-936(1986).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M13178; AAA66867.1; -.
DR   PIR; A24299; VCPV19.
DR   InterPro; IPR001403; Parvo_coat.
DR   Pfam; PF00740; Parvo_coat; 1.
KW   Coat protein; Glycoprotein.
FT   CARBOHYD     46     46       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    184    184       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    220    220       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    293    293       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   781 AA;  86015 MW;  8C6254DBD0576B07 CRC64;
```

```
  Query Match             3.4%;  Score 113.5;  DB 1;  Length 781;
  Best Local Similarity   19.5%;  Pred. No. 2.3;
  Matches  128;  Conservative   51;  Mismatches  191;  Indels  287;  Gaps   28;

Qy    112 GEYSKIPDGLVKFSVDAEIR----EIYNEECPVVTDVSVPLDGRQWSLSIFSFPMFRTAY 167
          |: |  |  |   |  || |     : :|:      ||| | |               | |
Db     85 GQLSDHPHALSSSSSHAEPRGENAVLSSEDLHKPGQVSVQLPG-------------TNY 130

Qy    168 VAVANVENKEMSLDVVNDLIEWLNNLADWRY--------------VVDSEQWINFTNDT 212
          |   |    |:        ::    : |:||               | | |   |  |:|
Db    131 VGPGN----ELQAGPPQSAVDSAARIHDFRYSQLAKLGINPYTHWTVADEELLKNIKNET 186

Qy    213 TYYVRIRVLRPTYDVPDPTEGLVRTVSDYRLTYKAITCEANMPTLVDQGFWIGGQYALTP 272
          :  ::                | ||  | |     |  |    ||          :
Db    187 GFQAQV-----------------VKDY-FTLKG----AAAPVAHFQG---------SL 213

Qy    273 TSLPQYDVSEAYALHTLTFARPSSAAALAFVWAGLPQGGTAPAGTPAWEQASSGGYLTWR 332
            :| |: || |   |      |  :| |   ||   || : :    | :
Db    214 PEVPAYNASEKYPSMT------SVNSAEASTGAG---GGGSNSVKSMWSE---------- 254

Qy    333 HNGTTFPAGSVSYVLPEGFALERYDP-------------------------------- 358
            | || | ||:      | :  |||
Db    255 --GATFSANSVTCTFSRQFLIP-YDPEHHYKVFSPAASSCHNASGKEAKVCTISPIMGYS 311

Qy    359 -------------------------NDGSWTDFASAGDTVTFRQVAVDEVVVTNNPAGGG 393
                                    | ||    |    |||  ::|| :|    :  |||
Db    312 TPWRYLDFNALNLFFSPLEFQHLIENYGS---IAPDALTVTISEIAVKDV---TDKTGGG 365

Qy    394 ------------------------------SAPTFTVRV--PPSNAYTNTVFRNTLLET 420
                                         ||   : |  ||  ||      ||
Db    366 VQVTDSTTGRLCMLVDHEYKYPYVLGQGQDTLAPELPIWVYFPPQYAYLTVGDVNT---- 421

Qy    421 RPSSRRLELPMPPADFGQTVANNPKIEQSLLKETLGCY-LVHSKMRNPVFQLTPASSFGA 479
                      | :: : |    |  |    | | ||      |||       :
Db    422 ----------------QGISGDSK---KLASEESAFYVLEHSS-----FQLLGTGGTAS 456

Qy    480 VSFN-------------------NPGYERTRDLPDYTGIRDSFDQNMSTAVAHFRSLSH 519
          :|:                    || |     :||  |                ||||:|
Db    457 MSYKFPPVPPENLEGCSQHFYEMYNPLYGSRLGVPDTLG-----------GDPKFRSLTH 505

Qy    520 -SCSIVTKTYQGWEGVTNVNTPFGQFAHAGLLKNEEILCLADDLATRLTGVYPA------ 572
           :|  : :     | :|:|  |  :: |  |    |       ||:: : |
Db    506 EDHAIQPQNFMPGPLVNSVSTKEGDSSNTGAGKALTGLSTGTSQNTRIS-LRPGPVSQPY 564

Qy    573 ----TDNFAAAVSAFAANMLSSVLKSEATSSIIKSVGETAVGAA-----QSGLAKLP 620
              || :   ::|                     | |:|  | |     | |: : |
Db    565 HHWDTDKYVTGINAI------------------SHGQTTYGNAEDKEYQQGVGRFP 602


RESULT   8
SACB_STRSL
ID   SACB_STRSL     STANDARD;      PRT;   969 AA.
AC   Q55242;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Levansucrase precursor (EC 2.4.1.10) (Beta-D-fructofuranosyl
DE   transferase) (Sucrose 6-fructosyl transferase).
GN   FTF.
OS   Streptococcus salivarius.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1304;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 25975;
RX   MEDLINE=93322332; PubMed=8331080;
RA   Rathsam C., Giffard P.M., Jacques N.A.;
RT   "The cell-bound fructosyltransferase of Streptococcus salivarius: the
RT   carboxyl terminus specifies attachment in a Streptococcus gordonii
RT   model system.";
RL   J. Bacteriol. 175:4520-4527(1993).
CC   -!- CATALYTIC ACTIVITY: SUCROSE + (2,6-BETA-D-FRUCTOSYL)(N) =
CC       GLUCOSE + (2,6-BETA-D-FRUCTOSYL)(N+1) (OTHER SUGARS CAN
CC       ALSO ACT AS FRUCTOSYL ACCEPTORS).
CC   -!- SUBCELLULAR LOCATION: SECRETED. CELL-WALL BOUND.
CC   -!- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L08445; AAA71925.1; -.
DR   InterPro; IPR003469; Glyco_hydro_68.
DR   Pfam; PF02435; Glyco_hydro_68; 1.
KW   Transferase; Glycosyltransferase; Signal; Cell wall.
FT   SIGNAL       1      ?       POTENTIAL.
FT   CHAIN        ?    969       LEVANSUCRASE.
SQ   SEQUENCE   969 AA;  103983 MW;  D389B5B32ACF735A CRC64;


  Query Match             3.3%;  Score 113;  DB 1;  Length 969;
  Best Local Similarity   21.0%;  Pred. No. 3.4;
  Matches   81;  Conservative   47;  Mismatches  141;  Indels  116;  Gaps   17;

Qy     97 KYLDPAGATESARAVGEYSKIPDGLVKFSVDAEIREIYNEECPVVTDVSVPLDGRQWSLS 156
          |  |  |  .:  |||:   :  | |  |:  : | :      ||   ||| | |  : |
Db    568 KSTDAEGTVAAREAVGD-DVVMLGFVSDSLRGKYRPLNGS--GVVLTASVPADWRTSTYS 624

Qy    157 IFSFP-------MFRTAYV----AVANVENK---------EMSLDVVNDLIEWLNNLADW 196
          :: |        :  |:|:     :|  |||         :|: |   :::  : |  ||
Db    625 YYAVPVEGSSDTLLVTSYMTNRGGIAGAENKSTWAPSFLIKMNADDTTEVLPKMTNQGDW 684

Qy    197 RYVVDSEQWINFTNDTTYYVRIRVLRPTYDVPDPTEGLVRTVSDYRLTYKAITCEANMPT 256
          :   ||  ::  :  :  :             | |       |: : | |::
Db    685 IWDKSSESLVHVGDQNSAKL------------PNE-------DFNVDYYAVS------- 717

Qy    257 LVDQGFWIGGQYALTPTSLPQYD----VSEAYALHTLTFARPSSAAALAFVWAGLPQGGT 312
                  | | | : |  |    ||||: : |:|                 :  |
Db    718 ----------GYGLKPHTYPTVDGSTGVSEAHGVLTVT----------------VKDGKD 751

Qy    313 APAGTPAWEQASSGGYLTWRHN--------GTTFPA-GSVSYVLPEGFALERYDPNDGSW 363
            |  |    : : |     |:          ||: ||  :       |      :|:  :
Db    752 KKADKPETPVSPTEG----NHSVDDKTNKPGTSKPADNNQPSADKEDKPTNPTNPDSPAR 807

Qy    364 TDFASAGDTVTFRQVAVD-EVVVTNNPAGGGS---------APTFTVRVPPS------NA 407
          | |   ||       : |  | |   |: | |         |      ||  :
Db    808 TPFPYYGDHSNDNNSSNDHHVAVPVKPSTGDSVGDRRPVAQAAEIATPVPKTIVATGPTV 867

Qy    408 YTNTVFRNTLLETRPSSRRLELPMP 432
           ||||   :: ||       | | |
Db    868 PTNTVKEESVTET-------EAPKP 885


RESULT   9
GUXB_CELFI
ID   GUXB_CELFI     STANDARD;      PRT;  1090 AA.
AC   P50899;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Exoglucanase B precursor (EC 3.2.1.91) (Exocellobiohydrolase B)
DE   (1,4-beta-cellobiohydrolase B) (CBP120).
GN   CBHB OR CENE.
OS   Cellulomonas fimi.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Micrococcineae; Cellulomonadaceae; Cellulomonas.
OX   NCBI_TaxID=1708;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 456-461.
```

```
RC   STRAIN=ATCC 484;
RX   MEDLINE=96003898; PubMed=7575482;
RA   Shen H., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;
RT   "Cellobiohydrolase B, a second exo-cellobiohydrolase from the
RT   cellulolytic bacterium Cellulomonas fimi.";
RL   Biochem. J. 311:67-74(1995).
RN   [2]
RP   SEQUENCE OF 54-75.
RX   MEDLINE=93209933; PubMed=8458833;
RA   Meinke A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;
RT   "Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase
RT   D (CenD), a family A beta-1,4-glucanase.";
RL   J. Bacteriol. 175:1910-1918(1993).
RN   [3]
RP   SEQUENCE OF 54-78.
RX   MEDLINE=94197708; PubMed=8147863;
RA   Shen H., Tomme P., Meinke A., Gilkes N.R., Kilburn D.G.,
RA   Warren R.A.J., Miller R.C. Jr.;
RT   "Stereochemical course of hydrolysis catalysed by Cellulomonas fimi
RT   CenE, a member of a new family of beta-1,4-glucanases.";
RL   Biochem. Biophys. Res. Commun. 199:1223-1228(1994).
CC   -!- FUNCTION: HYDROLYSE CELLOHEXAOSE TO A MIXTURE OF CELLOTETRAOSE,
CC       CELLOTRIOSE AND CELLOBIOSE, WITH ONLY A TRACE OF GLUCOSE. IT
CC       HYDROLYSED CELLOPENTAOSE TO CELLOTRIOSE AND CELLOBIOSE, AND
CC       CELLOTETRAOSE TO CELLOBIOSE, BUT IT DID NOT HYDROLYSE CELLOTRIOSE.
CC       HAS ALSO WEAK ENDOGLUCANASE ACTIVITY. HYDROLYSES GLUCOSIDIC BONDS
CC       WITH INVERSION OF ANOMERIC CONFIGURATION.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC       in cellulose and cellotetraose, releasing cellobiose from the non-
CC       reducing ends of the chains.
CC   -!- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC       (CBD).
CC   -!- SIMILARITY: BELONGS TO CELLULASE FAMILY L (FAMILY 48 OF GLYCOSYL
CC       HYDROLASES).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L38827; AAB00822.1; -.
DR   HSSP; P07986; 1EXG.
DR   InterPro; IPR001919; CBD_2.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR000556; Glyco_hydro_48.
DR   Pfam; PF00553; CBD_2; 1.
DR   Pfam; PF00041; fn3; 3.
DR   Pfam; PF02011; Glyco_hydro_48; 1.
DR   PRINTS; PR00014; FNTYPEIII.
DR   PRINTS; PR00844; GLHYDRLASE48.
DR   ProDom; PD011903; Glyco_hydro_48; 1.
DR   SMART; SM00060; FN3; 3.
DR   PROSITE; PS00561; CBD_BACTERIAL; 1.
KW   Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
FT   SIGNAL       1     33       POTENTIAL.
FT   PROPEP      34     53
FT   CHAIN       54   1090       EXOGLUCANASE B.
FT   DOMAIN      54    699       CATALYTIC (BY SIMILARITY).
FT   DOMAIN     700    785       FIBRONECTIN TYPE-III 1.
FT   DOMAIN     794    884       FIBRONECTIN TYPE-III 2.
FT   DOMAIN     891    978       FIBRONECTIN TYPE-III 3.
FT   DOMAIN     989   1090       CELLULOSE-BINDING (BY SIMILARITY).
FT   ACT_SITE   513    513       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID   990   1089       BY SIMILARITY.
SQ   SEQUENCE   1090 AA;  114829 MW;  046BB9D956F2F399 CRC64;


  Query Match             3.3%;  Score 112;  DB 1;  Length 1090;
  Best Local Similarity   22.2%;  Pred. No. 4.8;
  Matches  130;  Conservative   55;  Mismatches  219;  Indels  182;  Gaps   29;

Qy      39 TGRQVSPPDNFTAAA--------------QDLAQSLD-ANTVTFPANISSMPEFRNWAK 82
           ||:    || : |||                ||:  : | |  : | |  |     |: ||
Db     526 TGK----PDTWNAAAPTGNPGLTVEVTSYGQDVGVAADTARALLFYAAKSGDTASRDKAK 581

Qy      83 GKIDLDSDSIGWYFKYLDPAGATESARAVGEYSK-------------IPDG--------- 120
             :|       :    || | : :    |:| :              || |
Db     582 ALLD------AIWANNQDPLGVS-AVETRGDYKRFDDTYVANGDGIYIPSGWTGTMPNGD 634

Qy     121 LVKFSVD-AEIREIYNEECPVVTDVSVPLDGRQWSLSIFSFPMFR----TAYVAVANVEN 175
           ::|  |   :||  | :: |  : |   |||        : | ||     |  |||
Db     635 VIKPGVSFLDIRSFYKKD-PNWSKVQTFLDGG-------AEPQFRYHRFWAQTAVAGA-- 684

Qy     176 KEMSLDVVNDLIEWLNNLADWRYVVDSEQWINFTNDTTYYVRIRVLRPTYDVPDPTEGLV 235
                         |||:  : |       | |||         ||  ||   :  |
Db     685 ----------------LADYARLFDDG---TTTPDTT--------APT--VPTGLQAGV 714

Qy     236 RTVSDYRLTYKAITCEANMPTL-VDQGFWIGGQYALTPTSLPQYDVSEAYALHTLTF--- 291
           | ::  ::: | | :  :   | :|    |   |  :      | |||     |
Db     715 VTSTEATISWTASTDDTRVTGYDVYRGATKVGTATTTSFTDTGLTASTAYAYTVRAFDAA 774

Qy     292 ---ARPSSAAALAFVWAGLPQGGTAPAGTPAWEQASSGGYLTWRHNGTTFPAGSVSYVLP 348
             : ||  |||       |   |||:       :|:   :|   : :|  ||
Db     775 GNVSAPS--AALTVTTKATPSDTTAPSVPAITSSSSTANSVTIGWSASTDNAG------- 825

Qy     349 EGFALERYDPNDG----SWTDFASAGDTVTFRQVAVDEVVVTNNPAGGGSAPTFTVRVPP 404
           | | ||   |    : |   :  ||      | :  |   : ||  |||:  | |
Db     826 -GSGLAGYDVYRGATRVAQTTALTFTDTGLTASTAYEYTVRARDVAGNVSAPSTAVSV-- 882

Qy     405 SNAYTNTVFRNTLLETRPSSRRLELPMPPADFGQTVANNPKIEQSLLKET------LGCY 458
                 | :| | :     |  ||          :  :    |      |  |
Db     883 ----------TTKSDTTPDT---TAPSVPAGLAAMTVTETSVALTWNASTDTGGSGLKGY 929

Qy     459 LVHSKMRNPVFQLTPASSFGAVSFNNPG------YERTRDLPDYTGIRDSFDQNMSTAVA 512
           |:          |   |    |: : |      |: |    |  |       |:| |
Db     930 DVYR-------GATRVGSTTTASYTDTGLTAATAYQYTVRATDNAG-------NVSAA-- 973

Qy     513 HFRSLSHSCSIVTKTYQ------------GWEG-------VTNVNT 539
               | : |: ||| |             |         :||  |
Db     974 -----SAALSVTTKTPQTGGSCSVAYNASSWNSGFTASVRITNTGT 1014


RESULT  10
P1P_LACLC
ID   P1P_LACLC      STANDARD;      PRT;  1902 AA.
AC   P16271;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   PI-type proteinase precursor (EC 3.4.21.-) (Wall-associated serine
DE   proteinase).
GN   PRTP.
OS   Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OG   Plasmid pWV05.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Lactococcus.
OX   NCBI_TaxID=1359;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WG2;
RX   MEDLINE=88149035; PubMed=3278687;
RA   Kok J., Leenhouts K.J., Haandrikman A.J., Ledeboer A.M., Venema G.;
RT   "Nucleotide sequence of the cell wall proteinase gene of
RT   Streptococcus cremoris Wg2.";
RL   Appl. Environ. Microbiol. 54:231-238(1988).
CC   -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC       GROWTH OF THE BACTERIA ON MILK.
CC   -!- CATALYTIC ACTIVITY: ENDOPEPTIDASE ACTIVITY WITH VERY BROAD
CC       SPECIFICITY, ALTHOUGH SOME SUBSITE PREFERENCE HAVE BEEN NOTED,
CC       E.G. LARGE HYDROPHOBIC RESIDUES IN THE P1 AND P4 POSITIONS, AND
CC       PRO IN THE P2 POSITION. BEST KNOWN FOR ITS ACTION ON CASEINS,
```

```
CC       ALTHOUGH IT HAS BEEN SHOWN TO HYDROLYSE HEMOGLOBIN AND OXIDIZED
CC       INSULIN B-CHAIN.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC       SUBTILASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M24767; AAA17677.1; -.
DR   HSSP; P00782; 1S01.
DR   MEROPS; S08.019; -.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR003137; PA.
DR   InterPro; IPR000209; Peptidase_S8.
DR   Pfam; PF00746; Gram_pos_anchor; 1.
DR   Pfam; PF02225; PA; 1.
DR   Pfam; PF00082; Peptidase_S8; 3.
DR   PRINTS; PR00723; SUBTILISIN.
DR   PROSITE; PS00136; SUBTILASE_ASP; 1.
DR   PROSITE; PS00137; SUBTILASE_HIS; 1.
DR   PROSITE; PS00138; SUBTILASE_SER; 1.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; 1.
KW   Hydrolase; Serine protease; Cell wall; Zymogen; Signal; Plasmid;
KW   Transmembrane.
FT   SIGNAL        1     33
FT   PROPEP       34    187
FT   CHAIN       188   1902       PI-TYPE PROTEINASE.
FT   DOMAIN      188   1876       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1877   1895       MEMBRANE ANCHOR (POTENTIAL).
FT   DOMAIN     1896   1902       CYTOPLASMIC (POTENTIAL).
FT   ACT_SITE    217    217       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    281    281       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    620    620       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   DOMAIN     1867   1872       CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT                                PROTEINS.
SQ   SEQUENCE   1902 AA;  199910 MW;  2901C7F19B2E5D0B CRC64;


  Query Match             3.3%;  Score 111;  DB 1;  Length 1902;
  Best Local Similarity   21.5%;  Pred. No. 13;
  Matches  123;  Conservative   63;  Mismatches  210;  Indels  176;  Gaps   31;

Qy      38 RTGRQVSPPDNFTAAAQ---DLAQSLDANT-----VTFPA------NISSMPEFRNWAKG 83
           : |  :: |   ||  :    | :| |       : |         |:  |  | :|  |
Db     776 KAGSNITVPAGKTAQIEFTLSLPKSFDQQQFVEGFLNFKGSDGSRLNLPYMGFFGDWNDG 835

Qy      84 KIDLDSDSIGWYFKYLDPAGATESARAVGEYSKIP------------------------ 118
           ||    : | :      |||        | :  :|
Db     836 KIVDSLNGITY-----SPAG--------GNFGTVPLLTNKNTGTQYYGGMVTDADGNQTV 882

Qy     119 -DGLVKFSVDAEIREIYNEECPVVTDVSVPLDGRQWSLSIFSFPMFRTAYVAVANVENKE 177
            |  : || |     :||       |:|:                    |  : |: | :
Db     883 DDQAIAFSSDK--NALYN-------DISM------------------KYYLLRNISNVQ 914

Qy     178 MS-LDVVNDLIEWLN---NLADWRYVVDSEQWINFTN---DTTYYVRIRVLRPTYDVPDP 230
           :  ||   : :  |:   ||    |   |:|:| :     | |||          |
Db     915 VDILDGQGNKVTTLSSSTNLTKTYYNAHSQQYIYYNAPAWDGTYY-------------DQ 961

Qy     231 TEGLVRTVSDYRLTYK---------------AITCEANMPTL--------VDQGFWIGG 266
           :| ::|  |   ||:                   ::  ||:         : |
Db     962 RDGNIKTADDGSYTYRISGVPEGGDKRQVFDVPFKLDSKAPTVRHVALSAKTENG---KT 1018

Qy     267 QYALTPTSLPQYDVSEAYALHTL-TFARPSSAAALAFVWAGLPQGGTAPAGTP-AWEQAS 324
           || ||  :  : |:|   |  :: |     :     |  ||    |     || : |||
Db    1019 QYYLTAEA--KDDLSGLDATKSVKTAINEVTNLDATFTDAGTTADGYTKIETPLSDEQAQ 1076

Qy     325 SGG--------YLTWRHNGTTFPAGSVSYVLPEGFALERYDPNDGSWTDFASAGDTVTFR 376
           : |        |||   :  |   ||       | |     | |   |  |:  | |
Db    1077 ALGNGDNSAELYLTDNASNATDQDASVQKPGSTSFDL---IVNGGGIPDKISS--TTTGY 1131

Qy     377 QVAVDEVVVTNNPAGGGSAPTFTVRVPPS--NAYTNTVFR----NTLLETRPSSRRLELP 430
           :          |  |||:  ||:   | :    |||   :    ||  :   :|    :|
Db    1132 EA---------NTQGGGTY-TFSGTYPAAVDGTYTNAQGKKHDLNTTYDAATNSFTASMP 1181

Qy     431 MPPADF-GQTVANNPKIEQSLLKETLGCYLVHSKMRNPVFQLTPASSFGAVSFNNPGYER 489
           :  ||:  |      |    |||         :|:|     || | :|  : |||   :
Db    1182 VTNADYAAQVDLYADKAHTQLLKH------FDTKVR-----LT-APTFTDLKFNNGSDQT 1229

Qy     490 TRDLPDYTGI--RDSFDQNMSTAVAHFRSLSH 519
           :      ||    |:   |:   ||   :  |
Db    1230 SEATIKVTGTVSADTKTVNVGDTVAALDAQHH 1261


RESULT  11
GUNV_ERWCA
ID   GUNV_ERWCA     STANDARD;      PRT;   505 AA.
AC   Q47096;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Endoglucanase V precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase V)
DE   (Cellulase V).
GN   CELV.
OS   Erwinia carotovora.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Pectobacterium.
OX   NCBI_TaxID=554;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SCRI193;
RX   MEDLINE=94067016; PubMed=8246888;
RA   Cooper V.J.C., Salmond G.P.C.;
RT   "Molecular analysis of the major cellulase (CelV) of Erwinia
RT   carotovora: evidence for an evolutionary 'mix-and-match' of enzyme
RT   domains.";
RL   Mol. Gen. Genet. 241:341-350(1993).
CC   -!- FUNCTION: ENDOGLUCANASE WITH SOME EXOGLUCANASE ACTIVITY. THE PH
CC       OPTIMUM IS ABOUT 7.0 AND THE TEMPERATURE OPTIMUM ABOUT 42 DEGREES
CC       CELSIUS.
CC   -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC       linkages in cellulose.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC       HYDROLASES).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X76000; CAA53592.1; -.
DR   HSSP; O85465; 1A3H.
DR   InterPro; IPR001956; CBD_3.
DR   InterPro; IPR001547; Glyco_hydro_F5.
DR   Pfam; PF00942; CBD_3; 1.
DR   Pfam; PF00150; cellulase; 1.
DR   ProDom; PD001947; CBD_3; 1.
DR   PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW   Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT   SIGNAL        1     31       POTENTIAL.
FT   CHAIN        32    505       ENDOGLUCANASE V.
FT   DOMAIN       32    334       CATALYTIC.
FT   DOMAIN      335    352       LINKER.
FT   DOMAIN      353    505       CELLULOSE-BINDING (BY SIMILARITY).
FT   ACT_SITE    168    168       PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE    256    256       NUCLEOPHILE (BY SIMILARITY).
SQ   SEQUENCE   505 AA;  54900 MW;  DBEA9337BB4D2623 CRC64;
```

```
  Query Match             3.3%;  Score 110.5;  DB 1;  Length 505;
  Best Local Similarity   19.0%;  Pred. No. 2;
  Matches   87;  Conservative   69;  Mismatches  177;  Indels  125;  Gaps   24;

Qy     153 WSLSIFSFPMFRTA--YVAVANVENKEMSLDVVNDLIEWLNNLA-----DWRYVVDSEQW 205
           | :::|   |:  |  |::  :: ||     | : :    :|      ||  : |::
Db      85 WGINVFRVAMYTAADGYISNPSLANK------VKEAVAAAQSLGVYIIIDWHILSDNDPN 138

Qy     206 INFTNDTTYYVRIRVL-----RPTYDVPDPTEGLVR---TVSDYRLTYKAITCEANMP-- 255
           |      |::  :  |        |:: :   | |     :  | |     |  :  |
Db     139 IYKAQAKTFFAEMAGLYGSSPNVIYEIANEPNGGVTWNGQIRPYALEV-TDTIRSKDPDN 197

Qy     256 -TLVDQGFWIGGQYALTPTSLPQYDVSEAYALH-----------TLTFARPSSAAALAF 302
            :|  | |    :      ||  | :  ||||             : :|:   ||
Db     198 LIIVGTGTWSQDIHDAADNQLP--DPNTMYALHFYAGTHGQFLRDRIDYAQSRGAAIFVS 255

Qy     303 VWAGLPQGGTAPAGTPAWEQASSGGYL----TW----RHNGTTFPAGSVSYVLPEGFALE 354
           |      ||: |        : | :|    ||     : | ::   |::       ||
Db     256 EW------GTSDA------SGNGGPFLPESQTWIDFLNNRGVSWVNWSLTDKSEASAALA 303

Qy     355 RYDPNDGSWTD--FASAGDTVTFRQVAVDEVVVTNNPAGGGSAPTF-TVRVPPSNAYTNT 411
           | ||:   :::|  |       | :     |||  ||  |    | |  |
Bb     304 PGASKSGGWTEQNLSTSGKFVR-------EQIRAGANLGGGDTPTTPTEPTNPGNGTTGD 356

Qy     412 V---FRNTLLETRPSSRRLELPMPPADFGQTVANNPKIEQSLLK---------------- 452
           |   :||  ::  ||   : : :   : | |      |: | |:
Db     357 VVLQYRN--VDNNPSDDAIRMAVNIKNTGST-----PIKLSDLQVRYYFHDDGKPGANLF 409

Qy     453 ---ETLGCYLVHSKMRNPVFQLTPASSFGAVSFNN------PGYERTRDLP------DYT 497
              :|   : :    |      |: :  |:|::      || | | ::       |::
Db     410 VDWANVGPNNIVTSTGTPAASTDKANRYVLVTFSSGAGSLQPGAE-TGEVQVRIHAGDWS 468

Qy     498 GIRDSFDQNMSTAVAHFRSLSHSCSIVTKTYQGWEGVT 535
           : :: | :    |              :|  |: :|
Db     469 NVNETNDYSYGANVT-------------SYANWDKIT 492


RESULT  12
FLGE_CAUCR
ID   FLGE_CAUCR     STANDARD;      PRT;   591 AA.
AC   P35806;
DT   01-JUN-1994 (Rel. 29, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Flagellar hook protein flgE.
GN   FLGE OR FLAK OR CC0902.
OS   Caulobacter crescentus.
OC   Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC   Caulobacter.
OX   NCBI_TaxID=69394;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RA   Mullin D.A., Mullin A.H., Newton A.;
RT   "Organization and expression of Caulobacter crescentus genes needed
RT   for assembly and function of the flagellar hook.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=21173698; PubMed=11259647;
RA   Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA   Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA   Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA   Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA   Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA   Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT   "Complete genome sequence of Caulobacter crescentus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
RN   [3]
RP   SEQUENCE OF 1-52 FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=86089096; PubMed=4078896;
RA   Ohta N., Chen L.S., Swanson E., Newton A.;
RT   "Transcriptional regulation of a periodically controlled flagellar
RT   gene operon in Caulobacter crescentus.";
RL   J. Mol. Biol. 186:107-115(1985).
CC   -!- SIMILARITY: BELONGS TO THE FLAGELLA BASAL BODY ROD PROTEINS
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF072135; AAC33329.1; -.
DR   EMBL; AE005767; AAK22886.1; -.
DR   EMBL; M28551; AAA23042.1; -.
DR   TIGR; CC0902; -.
DR   InterPro; IPR001444; Flag_bb_rod.
DR   Pfam; PF00460; flg_bb_rod; 1.
DR   PROSITE; PS00588; FLAGELLA_BB_ROD; 1.
KW   Flagella; Complete proteome.
SQ   SEQUENCE   591 AA;  61207 MW;  FD4EEA319479A161 CRC64;


  Query Match             3.2%;  Score 109;  DB 1;  Length 591;
  Best Local Similarity   19.6%;  Pred. No. 3.2;
  Matches  130;  Conservative   76;  Mismatches  236;  Indels  220;  Gaps   31;

Qy      14 RRGTRNVRVSANTVTVNGRRNQRRRTGRQVSPPDNFTAAAQDLAQSLDAN---------- 63
           :| | |     :|:  :| :||    |   :    |  : | | ||  :|
Db      37 KRSTSNF----STLVTSGNKNQTYSAGGVKAQTHQF-ISQQGLTQSTTSNLDISISGAGF 91

Qy      64 --TVTFPANISSMPEFRNWAK-GKIDLDS-----DSIGWYFK-YL----------DPAGA 104
             |   | |:::  : |:: : |   ||:     :  | | : :|          ||:
Db      92 FVTTEKPENLTA-TDTRSFTRAGSFQLDNLGYLRNDAGLYLQGWLADPVSGLITPDPSDL 150

Qy     105 TESARA-VGEYSKIPDGLVKFSVDAEIREIYNEECPVVTDVSVPLDGRQWSLSIFSFPMF 163
           : |   ||      :   :  |:| :|     | ||   ||
Db     151 MQLASINVGSVGGTAEKTTRVGVNANLR----SEQPVAAAVSYK---------------- 190

Qy     164 RTAYVAVANVENKEMSLDVVNDLIEWLNNLADWRYVVDSEQWINFTNDTTYYVRIRVLRP 223
               |  |   :|                          ||||      ||
Db     191 ----VGTAGSPSK-------------------TNVVDSA-----TNS----------H 210

Qy     224 TYDVPDPTEGLVRTVSDYRLTYKAITCEANMPTLVDQGFWIGGQYALTPTSLPQYDVSEA 283
           |||   : |:   ||            |   |||    |     :  |    ||
Db     211 NYDVVYSSTGIANPVS----------GNNEYLVD----IKENGVIVATGKVAYD---- 250

Qy     284 YALHTLTFARPSSAAALAFVWAGLPQGGTAPAGTPAWEQASSGGYLTWRHN--GTTFPAG 341
                        ||    | : :   | :|          :|   | | |  |||
Db     251 ------------AATNELVSSTIDYKGASPV---------TGSMTTTRINAAGTTVNLA 288

Qy     342 SVSYVLPEG------FALERYDPNDGSWTDFASAGDTVTFRQVAVDEVVVTNNPAGGGSA 395
           :  |   |       | : |||:  | :|:|                          |
Db     289 DLGIVNASGADDAEVVAGKLYDPSTWSMSDYAKDNSK-------------------GVK 328

Qy     396 PTFTVRVPPSNAY--TNTVFRNTLLETRPSSRRLELPMPPADFGQTVANNPKIEQSLLKE 453
           | | |::| |::     ||  : |    |:    ||   | |       | :|   :::
Db     329 PDFEVQIPLSDSKGGQRTVTLSMLKGPGPNQWYAELRAKPGDLANN--GNGQISTGIIE- 385

Qy     454 TLGCYLVHSKMRNPVFQLTPASSFG-----AVSFNNPGYERTRDLP-----------DY 496
           :    |::|       | ||     |::  : ||      |            |
Db     386 ----FTTDGKLKN------TGSLFGTTSPTAITIKSSGYIAPTVTPPAVQPPTPPTWADA 435

Qy     497 TGIRDSFDQNMSTAVAHFRSLSHSCSIVTKTYQGWEGVTNVNTPFGQFAHAGLLKNEEIL 556
           || :   |    : |   :  :| |:|       : |    | ||   :  : :   :
Db     436 LGIDEQEVQIDLASAAGGLTQYNSQSVV-------QSVNTNGTAFGNLTNIEIDEGGYVS 488
```

```
Qy    557 CLADDLATRLTGVYPATDNFAAAVSAFA-ANMLSSV------LKSEATSSIIKSVGETAV 609
Db    489 AIFDNGVTRRIA--------QVAIATFSNPNGLKGVNGNAYRVTNESGTYSLKAPSQGGA 540

Qy    610 GA 611
Db    541 GA 542
```

```
RESULT  13
GUNN_ERWCA
ID   GUNN_ERWCA     STANDARD;      PRT;   444 AA.
AC   Q59394;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Endoglucanase N precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase N)
DE   (Cellulase N).
GN   CELN.
OS   Erwinia carotovora.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Pectobacterium.
OX   NCBI_TaxID=554;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATROSEPTICA FCBR C18;
RX   MEDLINE=98299944; PubMed=9636315;
RA   Olsen O., Thomsen K.K., Weber J., Duus J.O., Svendsen I., Wegener C.,
RA   von Wettstein D.;
RT   "Transplanting two unique beta-glucanase catalytic activities into
RT   one multienzyme, which forms glucose.";
RL   Biotechnology 14:71-76(1996).
CC   -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC       linkages in cellulose.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC       HYDROLASES).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L39788; AAC37033.1; -.
DR   HSSP; O85465; 1A3H.
DR   InterPro; IPR001956; CBD_3.
DR   InterPro; IPR001547; Glyco_hydro_F5.
DR   Pfam; PF00942; CBD_3; 1.
DR   Pfam; PF00150; cellulase; 1.
DR   ProDom; PD001947; CBD_3; 1.
DR   PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW   Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT   SIGNAL        1     31       POTENTIAL.
FT   CHAIN        32    444       ENDOGLUCANASE N.
FT   ACT_SITE    168    168       PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE    256    256       NUCLEOPHILE (BY SIMILARITY).
SQ   SEQUENCE   444 AA;  48300 MW;  FA7E4179004CBB43 CRC64;
```

```
  Query Match             3.2%;  Score 108.5;  DB 1;  Length 444;
  Best Local Similarity   20.9%;  Pred. No. 2.3;
  Matches   70;  Conservative   48;  Mismatches  138;  Indels   79;  Gaps   19;

Qy    153 WSLSIFSFPMFRTA--YVAVANVENKEMSLDVVNDLIEWLNNLA-----DWRYVVDSEQW 205
Db     85 WGINVFRVAMYTAADGYISNPSLANK------VKEAVAAAQSLGVYIIIDWHILSDNDPN 138

Qy    206 INFTNDTTYYVRIRVL-----RPTYDVPDPTEGLVR---TVSDYRLTYKAITCEANMP-- 255
Db    139 IYKAQAKTFFAEMAGLYGSSPNVIYEIANEPNGGVTWNGQIRPYALEV-TDTIRSKDPDN 197

Qy    256 -TLVDQGFWIGGQYALTPTSLPQYDVSEAYALH-----------TLTFARPSSAAALAF 302
Db    198 LIIVGTGTWSQDIHDAADNQLP--DPNTLYALHFYAGTHGQFLRDRIDYAQSRGAAIFVS 255

Qy    303 VWAGLPQGGTAPAGTPAWEQASSGGYL----TW----RHNGTTFPAGSVSYVLPEGFALE 354
Db    256 EW------GTSDA------SGNGGPFLPESQTWIDFLNNRGVSWVNWSLTDKSEASAALA 303

Qy    355 RYDPNDGSWTD--FASAGDTVTFRQVAVDEVVVTNNPAGGGSAPTFTVRVP-----PSNA 407
Db    304 PGASKSGGWTEQNLSTSGKFVR-------EQIRAGANLGGGDTPT-TPTTPTEPTNPGNG 355

Qy    408 YTNTV---FRNTLLETRPSSRRLELPMPPADFGQT 439
Db    356 TTGDVVLQYRN--VDNNPSDDAIRMAVNIKNTGST 388
```

```
RESULT  14
VP4_ROTH3
ID   VP4_ROTH3      STANDARD;      PRT;   775 AA.
AC   P39033;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Outer capsid protein VP4 (Hemagglutinin) (Outer layer protein VP4)
DE   [Contains: Outer capsid proteins VP5 and VP8].
GN   S4.
OS   Human rotavirus (serotype G3 / strain AU-1).
OC   Viruses; dsRNA viruses; Reoviridae; Rotavirus.
OX   NCBI_TaxID=39013;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92356070; PubMed=1322955;
RA   Isegawa Y., Nakagomi O., Nakagomi T., Ueda S.;
RT   "A VP4 sequence highly conserved in human rotavirus strain AU-1 and
RT   feline rotavirus strain FRV-1.";
RL   J. Gen. Virol. 73:1939-1946(1992).
CC   -!- SUBCELLULAR LOCATION: Outer capsid.
CC   -!- PTM: VP8 IS ONE OF TWO TRYPSIN CLEAVAGE PRODUCTS OF VP4; THE
CC       OTHER PRODUCT IS VP5.
CC   -!- SIMILARITY: BELONGS TO THE ROTAVIRUS VP4 PROTEIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D10970; BAA01747.1; -.
DR   PIR; JQ1638; JQ1638.
DR   InterPro; IPR000416; Cap_VP4.
DR   Pfam; PF00426; VP4; 1.
KW   Coat protein; Glycoprotein.
FT   CHAIN         1    775       OUTER CAPSID PROTEIN VP4.
FT   CHAIN         1    241       OUTER CAPSID PROTEIN VP8.
FT   CHAIN       248    775       OUTER CAPSID PROTEIN VP5.
FT   CARBOHYD     17     17       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     32     32       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     97     97       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    132    132       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    183    183       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    198    198       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    237    237       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    567    567       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    613    613       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   775 AA;  87092 MW;  4148D0C91A89B18A CRC64;
```

```
  Query Match             3.2%;  Score 108;  DB 1;  Length 775;
  Best Local Similarity   20.9%;  Pred. No. 5.6;
```

```
  Matches   67;  Conservative   51;  Mismatches  132;  Indels   70;  Gaps   14;

Qy     324 SSGGYLTWRHNGTTFPAGSVSYVLPEGFALERYDP-NDGS------WTDFASAGDTVTFR 376
           |  |   : :|| |          ||  ||: |::   : |      | |  :  : | |
Db     319 SVNGVNDFNYNGGT---------LPTDFAISRFEVIKENSYVYVDYWDDSQAFRNMVYVR 369

Qy     377 QVA--VDEVV-----------VTNNPAGGGSAPTFT-VRVPPSNAYTNTVFRNTLLETRP 422
            :|  :::||           | |:|   | | | |   |  |  ||: |  |:|
Db     370 SLAANLNDVVCSGGSYSFALPVGNHPVMSGGAVTLTSAGVTLSTQYTDYVSLNSL----- 424

Qy     423 SSRRLELPMPPADFGQTVANNPKIEQSLLKETLGCYLVHSKMRNPVFQLTPASSFGAVSF 482
             |  | :    |         | :: :    |   |:       :::    |  ::
Db     425 -RFRFRLAVSEPSF--------SISRTRMSGIYGLPAVNPNNNAEYYEIAGRFSLISLVL 475

Qy     483 NNPGYERTRDLPDYTGIRDSFDQNMSTAVAHFRSLSHSCSIVTKTYQGWEGVTNVNTPFG 542
           |  |:    : :   :|   :: :      | |||   ::        : :     |
Db     476 TNDDYQ--TPIANSVTVRQDLERQLGELREEFNSLSQEIAVS-------QLIDLATLPLD 526

Qy     543 QFAHAGLLKN--EEILCLADDLATRLTGVYPATDNFAAAVSAFAANMLSSVLKSEATSSI 600
           |:    :|:  | :  :  ::  |       | : | |:|   :||       ||| ||:
Db     527 MFSMFSGIKSTVEAVKSMTTNVMKRF-----KTSSLANAISDLTSNM------SEAASSV 575

Qy     601 ----IKSVGETAVGAAQSGL 616
               ::|||   :  |:  |
Db     576 RLTSVRSVGTVTLPRARVSL 595


RESULT  15
ALS1_CANAL
ID   ALS1_CANAL     STANDARD;      PRT;  1260 AA.
AC   P46590;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Agglutinin-like protein 1 precursor.
GN   ALS1.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 11651 / B792;
RX   MEDLINE=95272392; PubMed=7752895;
RA   Hoyer L.L., Scherer S., Shatzman A.R., Livi G.P.;
RT   "Candida albicans ALS1: domains related to a Saccharomyces cerevisiae
RT   sexual agglutinin separated by a repeating motif.";
RL   Mol. Microbiol. 15:39-54(1995).
CC   -!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.
CC   -!- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).
CC   -!- SIMILARITY: TO YEAST SAG1.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L25902; AAC41649.2; -.
KW   Cell adhesion; Glycoprotein; Repeat; Signal.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18   1260       AGGLUTININ-LIKE PROTEIN 1.
FT   DOMAIN      433    792       10 X 36 AA TANDEM REPEATS.
FT   REPEAT      433    468       1-1.
FT   REPEAT      469    504       1-2.
FT   REPEAT      505    540       1-3.
FT   REPEAT      541    576       1-4.
FT   REPEAT      577    612       1-5.
FT   REPEAT      613    648       1-6.
FT   REPEAT      649    684       1-7.
FT   REPEAT      685    720       1-8.
FT   REPEAT      721    756       1-9.
FT   REPEAT      757    792       1-10.
FT   DOMAIN      983   1152       2 X 26 AA APPROXIMATE REPEATS.
FT   REPEAT      983   1043       2-1.
FT   REPEAT     1092   1152       2-2.
FT   DOMAIN      399    404       POLY-THR.
FT   DOMAIN      408    418       POLY-THR.
FT   DOMAIN      450    455       POLY-THR.
FT   DOMAIN      486    491       POLY-THR.
FT   DOMAIN      522    527       POLY-THR.
FT   DOMAIN      558    563       POLY-THR.
FT   DOMAIN      594    599       POLY-THR.
FT   DOMAIN      630    635       POLY-THR.
FT   DOMAIN      666    671       POLY-THR.
FT   DOMAIN      702    707       POLY-THR.
FT   DOMAIN      738    743       POLY-THR.
FT   DOMAIN      774    779       POLY-THR.
FT   DOMAIN      874    877       POLY-SER.
FT   CARBOHYD    471    471       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    579    579       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    615    615       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    687    687       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    723    723       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    820    820       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    886    886       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    918    918       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    973    973       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1045   1045       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1068   1068       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1260 AA;  132641 MW;  763D1063A2354C24 CRC64;


  Query Match             3.2%;  Score 108;  DB 1;  Length 1260;
  Best Local Similarity   20.0%;  Pred. No. 11;
  Matches  110;  Conservative   54;  Mismatches  207;  Indels  178;  Gaps   24;

Qy     199 VVDSEQWI-NFTNDTTY--------YVRIRVLRPTYDVPDPTEGLVRTVSDYRLTYKAIT 249
           |  :| |  :|   ||          | ||      : |:||       |    |   :|
Db     474 VTTTEYWSQSFATTTTVTAPPGGTDSVIIR------EPPNPTVTTTEYWSQSFATTTTVT 527

Qy     250 C-----------EANMPTLVDQGFWIGGQYALTPT-------------SLPQYDV---- 280
                       |   ||:    :|    || | |                | : |
Db     528 APPGGTDSVIIREPPNPTVTTTEYW-SQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTE 586

Qy     281 --SEAYALHTLTFARP--------------------SSAAALAFVWAGLPQG------ 310
             |::||  |   | |                    | : |      | | |
Db     587 YWSQSYATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVII 646

Qy     311 ----GTAPAGTPAWEQASSGGYLTWRHNGTTFPAGSVSYVLPEGFALERYDPN-----DG 361
                 |  | |:    | |      | | |    |      | |  ||
Db     647 REPPNPTVTTTEYWSQS----YAT--TTTITAPPGETDTV------LIREPPNHTVTTTE 694

Qy     362 SWTDFASAGDTVTFRQVAVDEVVVTNNP--------------------AGGGSAPTFTV 400
           |:   :  |||      | |::  |                      |  |   |  :
Db     695 YWSQSYATTTTVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVII 754

Qy     401 RVPPSNAYTNTVFRNTLLETRPSSRRLELPMPPADFGQTVANNPKIEQSLLKETLGCYLV 460
           | ||:   | | : :    |  :     :  ||      :     | :|:    :
Db     755 REPPNPTVTTTEYWSQSFATTTT-----VTAPPGGTDTVI-----IYESMSSSKI----- 799

Qy     461 HSKMRNPVFQLTPASSFGAVSFNNPGY--ERTRDLPDYTGIRDSFDQNMSTAVAHFRSLS 518
           |   | :  : |       ||: | |    | ||  :    :|   |  |:::    |
Db     800 -STSSNDITSIIP-------SFSRPHYVNSTTSDLSTF----ESSSMNTPTSISSDGMLL 847

Qy     519 HSCSIVTKTYQGWEGVTNVNTPFGQF-AHAGLLKNEE-----ILCLADDLATRLTGVYPA 572
           | ::||::    | : :      :  : :|:: | :     |:      |: ::    :
Db     848 SSTTLVTESETTTESICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTASTMSDSLSS 907

Qy     573 TDNFAAAVS-------------------AFAANMLSSVLKSEATSSIIKSVGE------- 606
           ||  :|  |                    | |||  : |    | | || |
Db     908 TDGISATSSDNVSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTF 967

Qy     607 TAVGAAQSG 615
```

```
          |: |  |||
Db    968 TSNGDNQSG 976


Search completed: May 29, 2002, 11:40:45
Job time: 117 sec
```